ANALYSES OF THE DNA FLANKING THE HYPERVARIABLE MINISATELLITE MS31A (D7S21) IN THE MALAYSIAN POPULATION

YUSMIN BINTI MOHD. YUSUF

A dissertation submitted to the Institute of Postgraduate Studies and Research, University of Malaya, in partial fulfilment of the requirements for the Degree of Master of Philosophy

JUNE 1999

Institute of Postgraduate Studies and Research

University of Malaya

Kuala Lumpur

Abstract

In this study, variations within the 5' flanking DNA region of the minisatellite MS31 (D7S21) were investigated. DNA samples of 310 unrelated individuals from the Malaysian population (103 Malays, 106 Chinese and 101 Indians) were obtained and assayed for base substitutional polymorphisms at positions -4, -220, and -108 relative to the MS31 minisatellite. The variations at positions -4, -220, and -108 resulted in the presence or absence of the *Alu*I, *Hga*I, and *Psp*1406I restriction sites, respectively, and were detected by simple PCR-based RFLP assays. For each polymorphic site, the frequencies of the three possible genotypes (designated +/+, +/-, and -/-) in the three racial groups were calculated and tested for Hardy-Weinberg equilibrium distribution. In addition, the heterozygosity (**h**) and power of discrimation (Pd) values specific to each race were calculated.

Chi-square test results for all three sites did not show any significant deviation from those expected from Hardy-Weinberg equilibrium. The **h** values for the *Alu*I site polymorphism ranged between 0.319 and 0.377, whilst Pd values ranged between 0.482 and 0.551. Similarly, the **h** values for the *Hga*I site polymorphism, ranged from 0.42 and 0.50, whilst the Pd values ranged between 0.56 and 0.63. Finally, for the *Psp*1406I polymorphisms, the **h** values ranged between 0.411 and 0.494, whilst the Pd values ranged between 0.565 and 0.620.

The chi square 2x2 contingency test was performed to determine if any association existed between different polymorphic sites (*Hga*I-*Alu*I and *Hga*I-*Psp*1406I). Haplotype analysis was carried out to determine the phase of double

heterozygous individuals. Results from the contigency test indicated that there was significant association among the three polymorphic sites.

During the *Hga*I assay, seven samples of Malay and Indian origins produced irregular banding patterns that were not normally observed in the general population. The seven samples were sequenced. Analysis of the nucleotide sequences revealed a 12 bp deletion in all these samples, located from positions -230C to -241A.

Abstrak

Dalam kajian ini, variasi yang terdapat pada kawasan 5' yang mengapit DNA bagi minisatelit MS31 (D7S2I) diperiksa. Sampel DNA dari 310 individu yang tiada pertalian darah daripada populasi Malaysia (103 Melayu, 106 China dan 101 India) telah diperolehi dan dijalankan pemeriksaan untuk polimorphisma 'substitional' bes pada kedudukan -4, -220, dan -108, iaitu kedudukan yang relatif kepada minisatellite MS31. Variasi yang terhasil pada posisi-posisi ini adalah disebabkan kehadiran atau ketiadaan tapak pemotongan bagi enzim *Alu*I, *Hga*I dan *Psp*1406I melalui kaedah PCR yang ringkas. Bagi setiap tapak polimorphik, kekerapan bagi tiga kemungkinan genotype (dinamakan +/+, +/- dan -/-) dalam ketiga kumpulan kaum ini dikira dan diuji untuk keseimbangan taburan Hardy-Weinberg. Sebagai tambahan, heterozigositi (**h**) dan kuasa diskriminasi (Pd) juga dikira bagi setiap kaum.

Ujian chi-square menunjukkan ketiga-tiga tapak memberikan nilai yang bererti daripada nilai yang dijangkakan daripada keseimbangan Hardy-Weinberg. Nilai **h** bagi tapak polimorphisma *Alu*I berada di antara 0.319 dan 0.377, manakala nilai Pd adalah di antara 0.482 dan 0.551. Nilai yang hampir serupa diperolehi untuk tapak polimorphisma *Hga*I iaitu nilai **h** di antara 0.42 dan 0.50, nilai Pd, di antara 0.56 dan 0.63. Akhir sekali untuk tapak polimorphik *Psp*1406I, nilai **h** nya adalah di antara 0.411 dan 0.494, dan nilai Pd nya di antara 0.565 dan 0.602.

Ujian kontigensi 2x2 chi-square telah dijalankan untuk menentukan assosiasi yang wujud di antara tapak polimorphik yang berbeza (*Hga*I-*Alu*I dan *Hga*I-*Psp*1406I). Analisa haplotip pula telah dijalankan untuk menentukan fasa bagi individu yang mempunyai heterozigus double. Keputusan yang di dapati menunjukkan assosiasi yang bererti di antara ketiga-tiga tapak polimorphik itu.

Semasa kajian *Hga*I dijalankan, di dapati tujuh sampel menunjukkan corak fragmen yang berbeza di-gel dibandinkan dengan kawalan positifnya. Sampel yang

berasal daripada kaum India dan Melayu ini dijujukkan. Analisa menunjukkan terdapat delesi sebanyak 12 bp pada semua sampel, yang berkedudukan -230C hingga ke -241A.

Acknowledgement

I would like to express my deepest gratitute to my supervisors, Prof. Dr. Koh Chong Lek and Assoc. Prof. Dr. Sam Choon Kook for their invaluable advice, guidance and assistance throughout the course of my thesis. I am also greatly indebted to them for their time and patience in helping me.

I would also like to thank Miss Lim Moo Eng and Mr. Edmund for their assistance especially in collecting the blood sample.

To my ex-lab-mates and my present lab-mates, thank you for the support and friendship. I'll miss all of you.

Special thanks is dedicated to Professor Neva E Haites, and members of the DNA Lab, Medical Genetics, University of Aberdeen, for their advice and assistance during my stay in Aberdeen.

I would like to thank the University of Malaya for providing me the fellowship for my study, and the Ministry of Science, Technology and the Enviroment for the IRPA grant 03-07-04-115 to Prof. Koh Chong Lek, which supported this project.

Last but not least, special thanks is due to the members of my family especially my husband for the help, care, encouragement and understanding. Thank you all very much.

CHAPTER TWO: MATERIALS AND METHODS

Abbreviations

%	: percentage
(v/v)	: volume per volume
(w/v)	: weight per volume
<	: less than
>	: more than
A	: adenine
APS	: ammonium persulphate
ATP	: adenosine 5'-triphosphate
bp	: base pair
BPB	: bromophenol blue
BSA	: bovine serum albumin
C	: cytosine
cm	: centimetre
dATP	: deoxyadenosine 5'-triphosphate
dCTP	: deoxycytidine 5'-triphosphate
ddNTP	: dideoxyribonucleoside 5'-triphosphate
df	: degree of freedom
dGTP	: deoxyguanosine 5'-triphosphate
dH_2O	: distilled water
DNA	: deoxyribonucleic acid
dNTP	: deoxyribonucleoside 5'-triphosphate
dsDNA	: double-stranded DNA

dTTP	: deoxythymidine 5'-triphosphate
EDTA	: ethylenediaminetetraacetate
EtBr	: ethidium bromide
EtOH	: ethanol
g	: gramme
G	: guanine
h	: heterozygosity
HCl	: hydrochloric acid
hr	: hour
HVR	: hypervariable region
kb	: kilobase pairs
M	: molar
mA	: milliampere
Mb	: megabase pairs
mg	: milligramme
$MgCl_2$	: magnesium chloride
min	: minute
ml	: millilitre
mm	: millimetre
mM	: millimolar
MVR-PCR	: minisatellite variant repeat mapping by the polymerase chain reaction
NaCl	: sodium chloride or natrium chloride

NaOAc	: sodium acetate or natrium acetate
NH_4OAc	: ammonium acetate
OD	: optical density
p	: probability
PAGE	: polyacrylamide gel electrophoresis
Pd	: power of discrimination
psi	: pounds per square inch
RFLP	: restriction fragment length polymorphism
rpm	: revolutions per minute
sdH_2O	: sterile-distilled water
SDS	: sodium dodecyl sulphate
sec	: second
SSC	: sodium chloride-sodium citrate
ssDNA	: single-stranded DNA
T	: thymine
TBE	: Tris-borate EDTA
Tris	: Tris (hydroxymethyl) methylamine
U	: unit
uv	: ultraviolet
VNTR	: variable number of tandem repeats
°C	: degree Celsius
µg	: microgramme
µl	: microlitre

μm	: micrometre
μM	: micromolar
χ^2	: Chi square

List Of Figures

List Of Tables: